DETAILED ACTION

EXAMINER'S AMENDMENT

An examiner's amendment to the record appears below. Should the changes and/or additions be unacceptable to applicant, an amendment may be filed as provided by 37 CFR 1.312. To ensure consideration of such an amendment, it MUST be submitted no later than the payment of the issue fee.

Authorization for this examiner's amendment was given in a telephone interview with Viktor Simkovic (Reg 56,012) on 7/7/2010.

The Application is amended as follows:

Claims

1. (Currently Amended) A method perform by a computer system, the method comprising:

extracting, by one or more processors associated with the computer system, a set of uniform resource locators (URLs) from one document or from multiple documents ~~associated with a single web host~~ downloaded from a web site;

identifying, by the one or more processors associated with the computer system, a sub-string[[s]] occurring in ~~multiple URLs in~~ the set of URLs as a session identifier[[s]], based on [[a]] ~~particular~~ at least one of a plurality of rules and based on multiple occurrences of the sub-string[[s]] occurring in ~~multiple URLs of~~ the set of URLs;

generating, by the one or more processors associated with the computer system, a clean set of URLs, derived from the set of URLs, by removing the session identifier[[s]]; [[and]]

determining, by the one or more processors associated with the computer system, ~~when at least one particular~~ additional URLs ~~has~~ that have already been crawled based on a comparison of ~~the particular~~ a clean set of the additional URLs to the clean set of generated URLs; and

where the clean set of the additional URLs is generated by removing another session identifier, or the identified session identifier, of the additional URLs.

2-3. (Canceled)

4. (Currently Amended) The method of claim 1, where the comparison of the ~~particular~~ clean set of the additional URLs to the clean set of generated URLs ~~comprising~~ comprises:

calculating a first fingerprint value ~~for a particular one of~~ derive from the clean set of additional URLs and a second fingerprint value ~~for a particular~~ derive from ~~of~~ the clean set of generated URLs ~~URLS in the clean set of URLs~~, and where the comparison is based on a comparison of the first fingerprint value with the second fingerprint value ~~of the particular URL to the fingerprint values of the URLs in the clean set of URLs~~.

5. (Currently amended) The method of claim 1, where the at least one of plurality rules comprises:

determining that the sub-string[[s]] does not reference content.

6. (Canceled)

7. (Currently Amended) The method of claim 1, where the at least one of plurality rules comprises:

determining that the sub-string[[s]] contains characters consistent with a session identifier.

8. (Currently Amended) The method of claim 1, further comprising:

downloading content from the ~~particular URL~~ additional URLs when the ~~particular URL is~~ additional URLS are determined to not already have been crawled.

9. (Currently amended) The method of claim 1, further comprising:

storing information based on the clean set of URLs for use in later determining whether the additional URLs have already been extracted; and

storing the set of URLs, including embedded session identifiers, for use in later accessing the set of URLs.

10. (Currently Amended) [[The]] **A** method performed by a computer system, the method comprising:

~~receiving~~ downloading, by a communication interface associated with the computer system, one or more documents from a web site ~~a set of uniform resource locator (URLs);~~

extracting, by one or more processors associated with the computer system, a set of uniform resource locators (URLs) from the downloaded one or more documents;

~~analyzing, by one or more processor associated with the computer system, the set of URLs for sub-strings that are structured in a manner consistent with session identifiers; and further analyzing, by one or more processors associated with the computer system, the set of URLs to identify one of the sub-strings as corresponding to a session identifier based on multiple occurrences of the sub-string in the set of URLs.~~

identifying, by the one or more processor associated with the computer system, a sub-sting occurring in the extracted set of URLs as a session identifier, based on the sub-string having a structure consistent with session identifiers and based on multiple occurrences of the sub-string in the extracted set of URLs;

generating, by the one or more processors associated with the computer system, a clean set of URLs from the extracted set of URLs by removing the identified session identifier;

determining, by the one or more processors associated with the computer system, whether additional URLs have already been crawled based on a comparison of a clean set of the additional URLs to the generated clean set of URLs;

where the clean set of the additional URLs is generated by removing another session identifier, or the identified session identifier, of the additional URLs.

11.-12. (Canceled)

13. (Currently Amended) The method of claim 10, further comprising: ~~removing identified session identifiers from the set of URLs; and~~ storing the ~~set of URLs, with the removed session identifiers, as a~~ generated clean set of URLs.

14. (Currently Amended) The method of claim 13, further comprising:

adding a generated session identifier to ~~URLs in~~ each of the generated clean set of URLs.

15. (Currently Amended) A device comprising:

a memory to store instructions; and

a processor to execute the instructions to implement:

at least one fetch bot to download content on a network from a single web site ~~locations specified by uniform resource locators (URLs);~~

~~a content manager to:~~

extract URLs from the downloaded content ~~,and~~ ;

identify a sub-string as a session identifier[[s]] from the URLs extracted from the downloaded content based ~~, at least in part,~~ on at least one of a plurality rules and based on multiple occurrences of the sub-string in the extracted URLs ~~multiple occurrences of the session identifiers from a single web site~~; [[and]]

~~a URL manager to~~ create a clean set of ~~versions of the~~ URLs ~~extracted from the downloaded content~~ by removing the session identifier[[s]] from the extracted URLs; [[and]]

[[to]] store the clean set of URLs ~~versions of the URLs.~~; and

determine whether additional URLs have already been crawled based on a comparison of a clean set of the additional URLs to the created clean set of URLs, where the clean set of the additional URLs is generated by removing another session identifier, or the identified session identifier, from the additional URLs.

16. (Currently Amended) The device of claim 15, where the ~~content manager~~ processor is further to identify the sub-string as a session identifier[[s]] based on locating ~~sub-strings, within the URLs, that contain~~ characters consistent with a session identifier[[s]] in the URLs extracted from the downloaded content.

17. (Previously presented) The device of claim 15, further comprising:

a database to store the downloaded content.

18. (Currently Amended) The device of claim 15, where the ~~content manager~~ processor is further to determine ~~when~~ whether the additional URLs have previously been stored by comparing the clean set of the additional URLs to the stored clean set of URLs. ~~clean versions of the additional URLs to the stored clean versions of the URLs extracted from the downloaded content.~~

19. (Currently Amended) The device of claim 15, where the session identifier[[s]] ~~include~~ includes characters from the extracted URLs ~~extracted from the downloaded content~~ that do not reference content.

20. (Currently amended) A system comprising:

one or more server devices comprising one or more processors to:

download one or more documents from a web site;

extract a set of uniform resource locators (URLs) from the one or more documents downloaded from the website;

identify a sub-sting occurring in the set of URLs as session identifier, based on the sub-string including characters that are structured consistent with session identifiers and based on multiple occurrences of the sub-string in the set of URLs;

generate a clean set of URLs from the set of URLs by removing the identified sub-string;

determine whether additional URLs have already been crawled based on a comparison of a clean set of the additional URLs to the generated clean set of URLs; where the clean set of the additional URLs are generated by removing another session identifier, or the identified session identifier, of the additional URLs.

~~means for receiving a set of uniform locators (URLs);~~

~~means for analyzing the set of URLs for sub-strings that are structured in a manner consistent with session identifiers; and~~

~~means for further analyzing the set of URLs to identify one of the sub-strings as corresponding to a session identifier based on multiple occurrences of the sub-string in the set of URLs.~~

21-23. (Canceled)

24. (Currently Amended) The system of claim [[23]] 20, where the one or more processors are further ~~comprising~~ to:

~~means for adding~~ add a generated session identifier to each URL[[s]] in the generated clean set of URLs.

25. (Currently Amended) One or more memory devices that include programming instructions executed by one or more processors, where the instructions causes the one or more processors to: ~~the one or more memory devices including~~ :

~~one or more instructions to~~ extract a set of uniform resource locators (URLs) from one document or from multiple document associated with a single web host;

~~one or more instructions to~~ identify, in the set of URLs, ~~sub-strings~~ a sub-string as a session identifier based on the sub-string ~~that contain at least a particular number of characters or have~~ having at least a ~~particular~~ specified measure of randomness and based on multiple occurrences the sub-string in the extracted set of URLs; and

~~one or more instructions to further identify, in the identified sub-strings one of the sub-strings as corresponding to a session identifier based on multiple occurrences of the sub-string in the set of extracted URLs~~

generate a clean set of URLs from the extracted set of URLs by removing the identified session identifier;

determine, by the one or more processors associated with the computer system, additional URLs have already been crawled based on a comparison of a clean set of the additional URLs to

the clean set of URLs ; wherein the clean set of the additional URLs are generated by removing another session identifier, or the identified session identifier, of the additional URLs.

26-28. (Canceled)

29. (Currently Amended) The one or more memory devices of claim 25 [[28]], further causes the one or more processors to: ~~comprising:~~

~~one or more instructions to~~ add a generated session identifier to URLs in the clean set of URLs when the URLs are to be used to access a web document.

30. (Currently Amended) The method of claim 1, where the ~~particular~~ at least one of the plurality rules comprises:

determining that the sub-string[[s]] exhibits at least a ~~particular~~ specified measure of randomness.

31. (Currently Amended) The method of claim 10, ~~where analyzing the set of URLs for sub-strings that are structured in a manner consistent with session identifiers~~ identifying the sub-sting occurring in the extracted set of URLs as a session identifier includes identifying ~~sub-strings that have~~ the sub-string as having at least a ~~particular~~ specified measure of randomness.

32. (Currently Amended) The device of claim 15, where the processor is further to: identify[[ing]] the session identifier[[s]] from the extracted URLs ~~extracted from the downloaded~~

~~content is further~~ based on identifying that the sub-string[[s]] ~~that exhibit~~ exhibits at least a ~~particular~~ specified measure of randomness.

33. (Currently Amended) The system of claim 20, where the one or more processors are further to:

identify the sub-sting occurring in the set of URLs as a session identifier based on the sub-string having at least a specified measure of randomness. ~~where the means for analyzing the set of URLs for sub-strings that are structured in a manner consistent with session identifiers comprising means for identifying sub-strings that have at least a particular measure of randomness.~~

Reason for Allowance

The following is an examiner's statement of reasons for allowance:

1. Applicant's arguments, see pp 10-24, filed 9/14/2009, with respect to the rejection of claims under 35 U.S.C 103 are persuasive. Accordingly, the rejection has been withdrawn and the prior art of records do not teach or suggest claims 1, 10, and 15 filed on 9/14/2009.

2. Claims 1, 4, 5, 7-10, 13-20, 24, 25, 29-30 are allowed.

3. While the Examiner has thoroughly reviewed the claims and has not located any errors, Applicant is encouraged to independently verify that the claims contain no typographical errors and that all claim terms have proper antecedent basis.

Any comments considered necessary by applicant must be submitted no later than the payment of the issue fee and, to avoid processing delays, should preferably accompany the issue fee. Such submissions should be clearly labeled "Comments on Statement of Reasons for Allowance."

Any inquiry concerning this communication or earlier communications from the examiner should be directed to KAREN C. TANG whose telephone number is (571)272-3116. The examiner can normally be reached on M-F 7 - 5.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, John Follansbee can be reached on (571)272-3964. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

/Karen C Tang/
Primary Examiner, Art Unit 2451